

# Jason (Zhihang) Dong

Ph.D. Student | Looking for Research/Applied/Machine Learning Scientist Intern Opportunities

in [linkedin.com/in/zhihangdong](https://www.linkedin.com/in/zhihangdong) | [github.com/zdong1](https://github.com/zdong1) | +1 814 548 6383 | [zdong@uw.edu](mailto:zdong@uw.edu)  
📍 211 Savery Hall, Box 353340, Seattle, Washington, USA 98115 | <https://zdong.org>

I am most interested in **machine learning and data mining**, especially forecasting of high-dimensional, noisy, unstructured and temporal-spatio data, such as phone/sensor/text data, and translating them into meaningful social outcomes. Theoretically, I am fascinated about & **geometric methods for machine learning**. I have extensive project experience in **natural language processing** problems involving natural language understanding.

## Skills

**Notes :** Skills with **bold** mean 50+ hours experience (400+ for languages) & ranked with familiarity

**Computer Languages :** **Python, R, Java**, Scala, C(Statistics Libraries)

**Machine Learning Toolkits :** **Spark, Torch, xgBoost, caffè, Tensorflow, H2O**, Hadoop, CTNK, EMR, Keras

**Databases :** **Neo4j, Cassandra, MongoDB, postgre, mySQL, RDS, mySQL, Hive**, CouchDB, Lucene

**Natural Language Processing :** **Tika, NTLK, SpaCy, TextBlob**, Stanford Core NLP

**Softwares :** **ArcGIS, Stata, tableau**, GeoDa, SAS, Qt **System :** Git, AWS

## Work Experience

- September, 2019 | Microsoft Corporation, Applied Scientist Intern, Redmond, WA  
June, 2019 > More info to come...  
[Data Mining](#)
- December, 2018 | Amazon.com Inc., Applied Scientist Intern, Seattle, WA  
August, 2018 > Worked as NLP scientist on a natural language processing problem related to contextual sentiment extraction and product quality forecasting  
[NLP](#) [Machine Learning](#)
- August, 2018 | Deloitte Services LP., Senior Data Scientist Intern, Seattle, WA  
June, 2018 > Created a protocol for organizational network analysis on employee interaction data using causal inference, exponential random graph model (simulation) and a predictive modeling using RNN;  
> Built the idea of resume recommendation architecture 3.0 for consulting projects powered by deep learning with pipeline from resume reading to recommendation interface  
[NLP](#) [Social Network](#) [R](#) [AI](#)
- September, 2017 | Center for Studies in Demography and Ecology, Summer Research Assistant, Seattle, WA  
June, 2017 > Provided 400+ hours of statistical/programming supports to 5 different mini-projects;  
> Need-Based Assistance such as implementing GIS ID Matrix; identification algorithms of fake patient ID in the system with < 0.1% error rate and Viz Project such as **leaflet.js** on interactive mapping  
[R](#) [AWS](#) [d3](#) [leaflet](#) [Python](#) [SQL](#) [Demography](#) [Mapping](#)

## Research Projects

**Notes :** Titles are followed by relevant publication reference tags (sample links found there) and starting year.

- Current** | **Swiss Cell Data Record Project, [5], 2017**  
**June, 2017** > Managed 400G+ sensor data through Spark and cluster computing; Designed spatio-temporal models for complex human mobility; Implemented the topological model for improvement on activity space coverage; Modeling of interactions using tensor networks and nonparametric models  
[Geometric Methods](#) [Machine Learning](#) [Human-Computer Interaction](#)
- Current** | **Conversational Sentiment Modeling, [-], 2018**  
**January 2018** > Proposed an innovated model taking the temporal dependence and contextual dependence into consideration; the new sentiment model is motivated by the need to evaluate the sentiment represent by a series of text or a conversation  
[Natural Language Processing](#)

Current  
February 2018

### Causal Inference on Travel Behaviors with Small Perturbation, [7], 2018

- > Modeled the heterogeneous variables of bike rental and physical exercises activities and forecast future activities using ideas expanded from anchor regression

Space-Time Modeling

## Languages

English ● ● ● ● ●  
Chinese ● ● ● ● ●  
Japanese ● ● ○ ○ ○

## + Honors & Awards

- > **Clarence and Elissa M. Schrag Endowed Fellow**  
University of Washington, 2016 (2 per Ph.D. Cohort)
- > **UW Cloud Computing Credits Research Awards**  
University of Washington, 2016
- > **Undergraduate Research Travel Awards**  
Penn State University, 2014 - 2016

## Education

2021/6 (est.) Ph.D. — University of Washington, Seattle

**Research Areas** : Machine Learning, Natural Language Processing & Geometric Methods

**Advisor** : Prof. Adrian Dobra

2019/3 (est.) M.S. Statistics — University of Washington, Seattle

2016/05 B.A. — Pennsylvania State University, Sociology with Concurrent Majors and Minors in Statistics, Geography, Media Studies, Psychology (5 majors and 3 minors)

## “ Publications

- > [10] Zhihang Dong, W. Chen and X. Ge “Causal Inference on Bike Rentals using Anchor Regression” *ArXiv Preprint*
- > [9] Zhihang Dong, S. Song and A. Wang “Toward a More Interpretable and Reusable Data Science : Analyses of Million Jupyter Notebooks” *ArXiv Preprint*
- > [8] Zhihang Dong “Conversational Sentiment Models” Submitted to *The Web Conference 2019 (WWW 2019)*
- > [7] Tongshuang Wu, Zhihang Dong, S. Song and M. Zhang “Interactive Attention Model Explorer for NLP Tasks with Unbalanced Data Sizes” Submitted to *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)* [Sample]
- > [6] Zhihang Dong and Tongshuang Wu. “Benchmarking Open Source NoSQL Databases Performance on Text Queries”. [Sample]
- > [5] Zhihang Dong, A. Dobra and Y. Chen “Robust Activity Space Estimation using Sensor Trajectory Convergence”. [Sample] Submitted to *2018 IEEE International Conference on Data Mining (ICDM)*
- > [4] Zhihang Dong, Yen-Chi Chen and Adrian Dobra (2018). “Projecting the Short-term Population Mobility using Cell Data Records”. Accepted to *2018 Joint Statistical Meetings (JSM)*. [Presentation]
- > [3] — (2017). “Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling”. Accepted to *International Conference on Population Geography*.
- > [2] — (2016). “Theorizing Urban Neighborhoods : Mapping the Interneighborhood and Intranighborhood Networks and Criminogenic Factors on Street Crime Victimization”. Accepted to *American Society of Criminology Annual Meeting*.
- > [1] Dong, Zhihang (2016). “Modeling Age Homogeneity : Age Homogamy And Marital Happiness Over The Life Course”. Accepted to *American Sociological Association Annual Meeting (ASA)*.

## Course Works

- > 0. STAT 535 : **Statistical Learning** [Link]
- > 1. CSE 599i : **Online and Adaptive Learning** [Link]
- > 2. STAT 564 : **Bayesian Statistics**[Link]
- > 3. MATH 515 : **Optimization**[Link]
- > 4. CSE 547 : **Machine Learning for Big Data**[Link]
- > 5. EE 576 : **Computer Vision**[Link]
- > 6. CSE 599d : **Advanced NLP Methods** [Link]
- > 7. CSE 544 : **Database Management** [Link]
- > 8. CSE 512 : **Data Visualization** [Link]
- > 9. CSE 521 : **Design& Analysis of Algorithms** [Link]
- > 10.EE 595 : **Representation Learning** [Link]
- > 11.LING 575 : **Novel NLP Applications** [Link]
- > 12.CSE 599I2 : **Alg. thru Geometric Lens**[Link]

## + Labs & Working Groups

- > **Augmented Intelligence Seminar (AI Sem)**  
2018 – Current [Link]
- > **Geometric Data Analysis Reading Group**  
2018 – Current [Link]
- > **Working Group for Applied, Bayesian and Computational Statistics**  
2016 – Current [Link]
- > **Space-Time Reading Group**  
2016 – 2017
- > **Human Factors in GI Science Lab, PSU Geography**  
2014 – 2015 [Link]